

# 웹 페이지 구조 분석을 통한 효과적인 동영상 검색용 키워드 추출 방법

(An Effective Keyword Extraction Method Based on Web Page Structure Analysis for Video Retrieval in WWW)

이종원<sup>†</sup>    최기석<sup>\*\*</sup>    장주연<sup>\*\*</sup>    낭종호<sup>\*\*\*</sup>  
 (JongWon Lee)    (Giseok Choi)    (Juyeon Jang)    (Jongho Nang)

**요약** 본 논문에서는 웹 동영상 페이지의 구조를 바탕으로 하여, 웹 동영상의 관리 및 검색을 위한 주석용 키워드를 자동 추출하는 방법을 제안한다. 제안 방법은 웹 동영상 페이지의 구조를 포함한 동영상의 개수와 주변 텍스트 구성의 복잡도를 기준으로 4가지 타입으로 구분하고, 타입 별로 키워드를 추출하는 방법을 달리한다. 1,087개의 웹 동영상 페이지(2,462개의 동영상)를 바탕으로 실험한 결과에 의하면 본 논문에서 제안하는 방법은 기존 웹 이미지 검색 시스템을 위한 키워드 추출 방법보다 재현율 면에서 18%의 성능 향상을 보였다. 따라서 논문에서 제안하는 방법은 일반적인 웹 동영상 검색 시스템을 위한 키워드 추출에 널리 적용 될 수 있다.

키워드 : 정보검색, 웹 동영상 검색, 멀티미디어 정보검색, 동영상 키워드 추출

**Abstract** This paper proposes an effective keyword extraction method for the Web videos. The proposed method classifies the Web video pages in one of 4 types. As such, we analyzed the structure of the Web pages based on the number of videos and the layout of the Web pages. And then we applied the keyword extraction algorithm fit to each page type. The experiment with 1,087 Web pages that have total 2,462 videos showed that the recall of the proposed extraction method is 18% higher than ImagerRover[2]. So, the proposed method could be used to build a powerful video search system for WWW.

**Key words** : Information Retrieval, Web video retrieval, Multimedia Information Retrieval, Keyword extraction for video

## 1. 서론

멀티미디어 데이터 처리기술의 발달은 멀티미디어 데

이터의 관리 및 검색을 중요하게 만들었으며, 그 중에서도 WWW에 산재된 데이터들의 증가로 인해서 웹 기반의 데이터 관리가 중요해지고 있다. 웹 기반의 멀티미디어 데이터 중에서 웹 이미지 관리 및 검색 시스템에 대해서는 활발한 연구가 진행되어 왔는데, 웹 페이지로부터 이미지와 관련된 키워드를 추출하기 위해서 텍스트 정보의 출현 빈도수와 태그 위치 등을 고려한 키워드 선정 방법[1-3]에 대해 연구되어 왔다. 그러나 이미지에 비해서 동영상은 파일 크기가 매우 크고, 웹에 많이 분포해 있지 않았을 뿐만 아니라, 몇 개의 특정 동영상 제공 업체에서 보유하는 동영상들이 대부분을 차지하여왔기 때문에 관련 연구가 활발하게 진행되지 않았다. 그러나 최근 들어 이전보다 더 많은 웹 사이트에서 동영상 콘텐츠를 제공하고 있고, UCC처럼 일반 사용자들도 쉽게 동영상을 웹 페이지에 게시할 수 있게 됨에 따라 웹 기반 동영상의 효율적인 검색기능이 요구 되고 있다.

본 논문에서는 웹 동영상 페이지를 대상으로 하여, 웹

· 이 논문은 2007 한국컴퓨터종합학술대회에서 '웹 페이지 구조 분석을 통한 효과적인 동영상 검색용 키워드 추출 방법'의 제목으로 발표된 논문을 확장한 것임

† 정 회 원 : 청강문화산업대학 e스포츠게임과 교수  
 jongwon@ck.ac.kr

\*\* 학생회원 : 서강대학교 컴퓨터학과  
 brix@mlneptune.sogang.ac.kr  
 jyjang@mlneptune.sogang.ac.kr

\*\*\* 종신회원 : 서강대학교 컴퓨터학과 교수  
 jhnang@sogang.ac.kr  
 논문접수 : 2007년 9월 27일  
 심사완료 : 2008년 1월 23일

Copyright ©2008 한국정보과학회 : 개인 목적이거나 교육 목적인 경우, 이 저작물의 전체 또는 일부에 대한 복사본 혹은 디지털 사본의 제작을 허가합니다. 이때, 사본은 상업적 수단으로 사용할 수 없으며 첫 페이지에 본 문구와 출처를 반드시 명시해야 합니다. 이 외의 목적으로 복제, 배포, 출판, 전송 등 모든 유형의 사용행위를 하는 경우에 대하여는 사전에 허가를 얻고 비용을 지불해야 합니다. 정보과학회논문지 : 시스템 및 이온 제35권 제3호(2008.4)

페이지로부터 동영상 관련 키워드를 추출하는 새로운 방법을 제안하고, 실험을 통해 결과를 분석한다. 웹 동영상 페이지와 웹 이미지 페이지는 서로 다른 특성을 가졌는데, 웹 이미지의 경우 이미지가 주변 텍스트를 설명하기 위한 보조 수단인데 반해서 웹 동영상은 주변 텍스트가 동영상을 설명하기 위한 보조 수단이 되는 차이점이 있다. 또한 웹 동영상의 경우, 거의 모든 웹 페이지에 비슷하게 분포되어 있는 웹 이미지와는 다르게 웹 동영상 페이지들은 특성 별로 분류가 가능하다. 이러한 특성을 바탕으로 제안한 방법에서는 웹 동영상 페이지로부터 키워드를 추출하기 위해, 먼저 웹 동영상 페이지를 구성 방식에 영향을 주는 요소(웹 페이지가 포함하는 동영상 수, 동영상으로부터 분포된 텍스트 형식)에 따라 4가지 타입으로 분류한다. 타입 별로 동영상 주변의 텍스트 블록을 찾고, 그 블록들로부터 타입 특성에 맞는 키워드 추출 알고리즘을 적용하여, 관련된 키워드를 추출하게 된다. 제안한 방법에 대한 성능 측정을 위하여, 웹 동영상 페이지로부터 키워드를 추출하는 시스템을 구현하고, 실험을 통하여 성능을 분석하고 평가하였다. 그 결과, 임의의 120개의 웹 동영상 페이지들에 대하여 제안한 방법이 기존 웹 이미지 시스템에서의 키워드 추출 방법인 ImageRover[2]를 웹 동영상에 적용한 결과보다 18%의 재현율 향상을 보였다. 타입 별로는 TYPE1을 제외한 모든 타입에 대하여 성능 향상을 보였고, 특히 복잡한 구조의 페이지인 TYPE2, TYPE4에서는 더 큰 성능 향상을 보였다. WWW에 분포한 일반적인 웹 동영상들의 약80%(2,462개의 동영상을 대상으로 조사)가 TYPE2,3,4에 포함되므로 제안한 방법에 대한 키워드 추출 방법은 일반적인 웹 동영상 검색 시스템에 널리 활용 될 수 있을 것으로 예상된다.

## 2. 기존 연구 분석

멀티미디어 데이터의 검색을 위한 키워드 추출방법에 대한 기존의 연구를 살펴본다.

### 2.1 일반적인 동영상 주석 방법에 관한 연구

동영상 주석 방법은 크게 두 가지 방법으로 나뉜다. 첫 번째 방법은 사람에게 의해 수동으로 주석을 다는 것이고, 다른 방법은 시스템에 의해서 자동으로 동영상에 주석을 다는 방식이다. VAnnotator[4]는 간단한 형태의 기본적인 윈도우 동영상 주석 도구이다. 저작자는 동영상을 재생하면서, 추가적으로 타임 라인을 움직여가며 주석을 다는 것이 가능한데, 이때 주석의 대상이 되는 단어는 사용자가 직접 결정하여 입력하게 한다. EVA[5] 역시 수동 주석 툴로써 웹 기반으로 구현되었다. 수동 주석 방법의 경우 정확도가 높은 장점이 있지만, 노동 집약적이고 시간이 많이 소요된다. 따라서 웹에 산재한

동영상 데이터를 대상으로 수동 주석 방법을 적용 시키는 것은 불가능하다. M. Bertini[6]의 방법은 학습 과정에서 동영상의 시각정보를 가지고 클러스터링 한 결과 그룹을 가지고 온톨로지를 구성하고, 후에 입력 동영상에 대한 비교 대상으로 사용한다. 주석이 자동으로 이루어지기는 하지만, 축구 동영상이라는 특정 타깃에 대하여 적용을 하였다. 이와 같이 기존의 동영상 주석 방법에 관한 연구들은 수동 주석 방법을 기반으로 하거나 도메인 한정적인 동영상에 대하여 자동 주석 방법의 연구가 진행되어 왔다. 그러나 웹 동영상은 무한하고, 다양한 장르를 포함하기 때문에 기존의 일반적인 동영상의 주석 방법들을 적용하는 것은 불가능하다.

### 2.2 웹 기반 이미지에 대한 키워드 추출 방법

웹 이미지 검색을 위해서는 먼저 검색 대상이 되는 이미지의 수집 과정에서 웹 페이지로부터 이미지를 추출하고, 주변 텍스트에 등장하는 단어들의 출현 빈도와 HTML 태그의 중요도, 그리고 이미지 파일 명과 같은 추가 정보들을 이용하여 주석에 사용될 키워드 들을 추출하는 자동 이미지 주석방법[7,8]이 연구되어 왔다.

WebSeer[1]는 텍스트와 HTML 소스 코드로부터 얻은 정보와 이미지 내용 정보로부터 얻은 결과로부터 주석용 키워드를 결정한다. 이미지 파일명, 이미지 캡션, ALT태그, HTML 제목, 하이퍼링크 등을 이용해서 단어를 추출하고, 각 단어들의 추출 태그에 따라 가중치를 부여하는 방식이다. ImageRover[2]는 WebSeer의 방법에 바탕을 두고 있으며, 해당 및 볼드, 이탤릭과 같이 이미지 내용을 직접적으로 표현 할 수 있는 태그나 시각적으로 강조된 텍스트 등에 대해 상대적으로 높은 가중치를 부여하며, 텍스트의 빈도수 및 이미지로부터 떨어진 거리를 고려하여 가중치를 부여한다. AMORE[3]는 테이블에서의 이미지 태그와 텍스트의 배치 패턴, 이미지 파일명과의 유사한 문자 개수 등과 같은 정보를 추가적으로 이용한다는 데에서 앞의 방법들과 차이가 있다. 이러한 방법들은 웹 이미지 검색에 관한 연구이지만, 동영상도 이미지의 연장으로 볼 수 있으므로, 웹 이미지 검색 방법을 그대로 적용함으로써 관련 키워드를 추출 할 수 있다. 그러나 웹 페이지 내에서 동영상이 차지하는 의미와 주변 텍스트와의 관계는 이미지에서와는 다르기 때문에, 기존의 방법을 그대로 적용하는 것은 적합하지 않고 웹 동영상 페이지의 특성에 맞는 키워드 추출 알고리즘이 필요하다.

## 3. 웹 동영상 페이지의 특징 분석 및 새로운 키워드 추출 알고리즘

본 장에서는 웹 동영상 페이지와 웹 이미지 페이지의 구체적인 차이점을 분석하고, 웹 동영상 페이지에 대한

특성을 파악한 후, 그 특성에 맞는 키워드 추출 알고리즘을 설명한다.

3.1 웹 동영상 페이지와 웹 이미지 페이지의 차이점

동영상과 이미지는 단순히 파일 포맷에서의 차이뿐만 아니라, 웹 페이지 내에서 미디어 데이터가 차지하는 의미가 다르기 때문에 같은 주석 방법을 사용하는 것은 적합하지 않다.

그림 1의 웹 페이지에는 뉴스기사에 대한 대표 이미지가 왼쪽 상단에 위치하고, 관련 동영상이 왼쪽 하단에 위치한다. 그림 1(a)에는 이미지와 연관이 있는 단어들에 표시를 하였고, 그림 1(b)에는 동영상과 관련이 있는 단어들에 표시를 하였다. 먼저 그림 1(a)에서는 소수의 단어들만 이미지와 상관관계를 가지고 있다. military, attack, guard, soldier라는 단어들은 이미지내의 객체를 잘 표현하고 있으나, 그런 단어들 외에도, Iraq, U.S., Education, Explosion 등과 같이 이미지만으로는 연상이 불가능한 단어들을 많이 포함하고 있는데, 이것은 이미지라는 데이터가 웹 페이지 내에서 의미적으로 중심이 되는 것이 아니라 텍스트를 위한 보조 수단의 역할을 하기 때문이라고 해석 할 수 있다. 그림 1(b)의 경우에는 동영상이 웹 페이지 내 단어들의 상당수가 연관성이 있는데, 이는 동영상이 텍스트에 포함된 짧은 내용을 뉴스의 형태로 설명하기 때문이라고 볼 수 있다. Baghdad, bombing 등과 같은 단어는 이미지로부터는 연상할 수 없지만 동영상은 관련 의미를 충분히 언급하고 있다. 또한 동영상의 경우 텍스트를 읽지 않더라도 동영상만으

로 웹 페이지의 내용을 모두 파악 할 수 있는데, 이는 동영상이 웹 페이지 내에서 중심이 되고, 주변 텍스트들이 동영상 데이터를 기술하기 위한 보조 수단이 된다는 것을 의미 한다. 이와 같이 웹 동영상 페이지는 웹 이미지 페이지와 다른 특성을 가지고 있으므로, 키워드 추출에 있어서 기존과는 다른 방법이 요구된다.

3.2 웹 동영상 페이지의 특징분석에 따른 타입 분류

키워드 추출의 대상이 되는 텍스트를 선정하는 데에 있어서 웹 페이지 내 포함된 동영상의 개수가 중요하다. 동영상 수가 하나일 경우 주변 텍스트들은 대부분 그 동영상에 대한 정보를 담고 있으나, 여러 개를 포함할 경우 텍스트들은 각기 다른 동영상을 기술할 수도 있고, 공통의 텍스트를 기술 할 수도 있다. 이와 같이 웹 동영상 페이지의 구조적인 특성에 의해 키워드를 추출하는 방법이 달라 질 수 있다. 다음에서는 웹 동영상 페이지 구조의 분류 기준이 될 수 있는 특징들에 대해서 논하고, 그 특징을 바탕으로 하여 동영상 페이지의 타입을 도출한다.

3.2.1 웹 동영상 페이지 구조의 분류 기준

(1) 웹 페이지 내 포함된 동영상과 주변 텍스트 분포  
 웹 페이지로부터 보다 정확한 키워드를 추출하기 위해서는 웹 페이지 내에서 동영상과 상대적으로 상관관계가 높은 텍스트 영역을 찾고, 그 영역으로부터 키워드를 추출해야 한다. 따라서 동영상과의 상관관계가 높은 텍스트 영역을 찾으려면, 먼저 웹 페이지에 있는 텍스트 영역을 분리하여 동영상과의 레이아웃 거리(Layout

**100 Americans die in Iraq during October**

By CHRISTOPHER BODEEN, Associated Press

BAGHDAD, Iraq - At least 78 people were 33 bombing victims of an attack on laborers Shiite slum. The U.S. military announced to combat this month.

Gunmen also killed hard-line Sunni academic Essam al-Rawi, head of the University Professors Union, as he was leaving home. At least 156 university professors have been killed since the war began. Hundreds, possibly thousands, more are believed to have fled to neighboring countries, although Education Ministry spokesman Basil al-Khatib said he had no specific numbers on those who have left the country.

Reuters Photo: A U.S. soldier from 172nd Stryker Brigade Combat Team stands guard on a road in...

Slideshow: Iraq

RELATED VIDEO

Blast hits Baghdad's Sadr City  
 A blast at Sadr City Marketplace kills at least 25 people in Baghdad. CNN's Awra Damon reports (October 30)  
 Play video | More CNN video

Sadr City, is a stronghold of the Mahdi Army has been the scene of repeated bomb attacks

(a) 웹 이미지와 관련 단어

**100 Americans die in Iraq during October**

By CHRISTOPHER BODEEN, Associated Press

BAGHDAD, Iraq - At least 78 people were 33 bombing victims of an attack on laborers Shiite slum. The U.S. military announced to combat this month.

Gunmen also killed hard-line Sunni academic Essam al-Rawi, head of the University Professors Union, as he was leaving home. At least 156 university professors have been killed since the war began. Hundreds, possibly thousands, more are believed to have fled to neighboring countries, although Education Ministry spokesman Basil al-Khatib said he had no specific numbers on those who have left the country.

Reuters Photo: A U.S. soldier from 172nd Stryker Brigade Combat Team stands guard on a road in...

Slideshow: Iraq

RELATED VIDEO

Blast hits Baghdad's Sadr City  
 A blast at Sadr City Marketplace kills at least 25 people in Baghdad. CNN's Awra Damon reports (October 30)  
 Play video | More CNN video

Sadr City, is a stronghold of the Mahdi Army has been the scene of repeated bomb attacks

(b) 웹 동영상과 관련 단어

그림 1 웹에서 텍스트와 미디어컨텐츠의 상관관계

distance)를 측정할 필요가 있다.

웹 동영상 페이지에서 동영상과 텍스트 영역의 레이아웃 거리를 계산하기 위해서 VIPS[9] DOM 트리를 이용한다. 웹 페이지의 VIPS DOM 트리에서 동영상노드로부터 폭 방향으로 멀리 떨어져 있을수록 화면상의 거리가 멀어지므로 레이아웃 거리는 커지게 된다. 그러나 DOM트리에서 깊이가 깊어지면 화면상의 배치가 세밀하고 복잡해질 뿐, 거리가 멀어지는 것이 아니므로 레이아웃 거리가 커지진 않는다.

동영상노드를  $N_v$ 라고 하고, 텍스트 노드를  $N_t$ 라고 할 때,  $N_v$ 에서  $N_t$ 까지의 거리( $d(N_v, N_t)$ )는 식 (1)과 같이 계산한다. 여기에서 집합  $P = \{p_0, p_1, \dots, p_n\}$ 은  $N_v$ 의 조상노드들의 집합이고, 집합  $Q = \{q_0, q_1, \dots, q_m\}$ 은  $N_t$ 의 조상노드들의 집합이라고 할 때,  $N_v$ 의 조상노드( $p_i$ )와  $N_t$ 의 조상노드( $q_j$ )는 트리구조의 어느 지점에서 같은 노드로 만나게 되는데, 이 때  $N_v$ 노드와  $p_i(q_j)$ 노드의 레벨차이로 레이아웃 거리가 계산된다. 식 (2)의  $L(N_i)$ 는 루트노드의 레벨을 0으로 했을 때 각 노드의 레벨을 계산한다.

$$d(N_v, N_t) = L(N_v) - L(p_i) - 1 \quad (1)$$

$$L(N_i) = \begin{cases} 0, & \text{if } N_i \text{ is the root node} \\ L(N_k) + 1, & \text{if } N_k \text{ is the parent node of } N_i \end{cases} \quad (2)$$

예를 들어 DOM트리구조가 그림 2와 같을 경우, 그룹A에 속한 텍스트노드( $N_t$ )들과 동영상노드( $N_v$ )의 공통 조상노드는  $P_0$ 노드이므로  $L(N_v) - L(P_0) - 1$ 과 같이 계산하면 거리가 0이 된다. 마찬가지로 그룹B에 속한 텍스트노드와 동영상노드는  $P_2$ 노드가 공통 조상노드이므로 거리는 2가 된다.

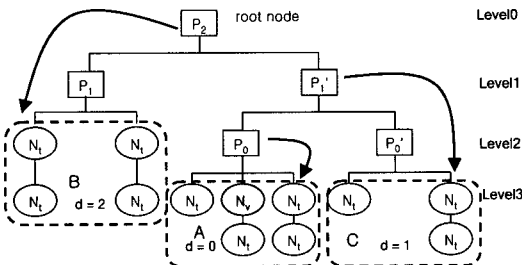


그림 2 거리계산 예제

(2) 웹 동영상 페이지가 포함하는 동영상의 수

웹 페이지 내 동영상이 한 개 위치해 있을 경우, 주변 텍스트들은 대부분 그 동영상에 관련한 내용을 기술하게 되지만, 여러 개 있을 경우 텍스트마다 다른 동영상을 설명할 수도 있고, 다수의 동영상을 공통으로 설명할 수도 있다. 웹 페이지 내에 있는 동영상 수에 따라서 텍스트가 동영상과 상관관계가 있을 수도 있고, 그렇지 않

을 수도 있기 때문에 웹 페이지가 포함하는 동영상의 수는 키워드 추출에 있어서 중요한 분류 기준이 된다.

### 3.2.2 웹 동영상 페이지의 구조에 따른 타입 분류

앞에서 설명한 웹 동영상 페이지의 구조에 대한 2가지 특성(동영상 수, 주변 텍스트 분포)을 기준으로 웹 동영상 페이지의 타입을 표 1과 같이 4가지로 분류할 수 있다.

표 1 웹 동영상 타입 분포(T=1)

타입	동영상수	텍스트분포	웹동영상분포
TYPE1	n = 1	$d \leq T$	16.3%
TYPE2		$d > T$	25.0%
TYPE3	n > 1	$d \leq T$	28.8%
TYPE4		$d > T$	29.9%

TYPE1은 웹 페이지가 단일 동영상을 포함하고, 동영상 주변의 텍스트가 비교적 간단한 타입이고, TYPE2는 단일 동영상만을 포함하면서 동영상 주변의 텍스트가 복잡한 구조를 가지는 형태이다. 여기에서 텍스트의 복잡도는 HTML페이지의 DOM트리에서 동영상 노드를 기준으로 주변 텍스트 노드들이 어느 정도의 레이아웃 거리를 가지고 분포했는지를 기준으로 판단한다. TYPE3은 웹 페이지에 여러 동영상이 포함되어 있고 텍스트들의 분포가 간단한 형태이며, TYPE4는 여러 동영상을 포함하면서 텍스트가 복잡하게 분포한 형태이다. 표 1에서 n은 웹 페이지가 포함하는 동영상의 개수를 의미하고, d는 DOM트리에서 동영상 노드로부터 텍스트 블록까지의 최대 레이아웃 거리를 의미하며, T는 텍스트 분포의 복잡한 정도를 결정하는 임계치값을 의미한다. 직접 수집한 1,087개의 웹 동영상 페이지(2,462개의 동영상을 포함)를 대상으로 분석한 결과, 주변에 텍스트가 전혀 없는 1.01%를 제외한 나머지 웹 동영상 페이지에 대해서, T가 1일 때 표 1과 같은 분포를 나타냈다. 동영상의 분포가 TYPE별로 유사한 수치를 기록하는 것으로 보아서, 이와 같은 방법으로 타입을 분류하는 것은 의미가 있다.

### 3.3 웹 동영상 검색을 위한 키워드 추출 알고리즘

#### 3.3.1 키워드 추출과 가중치 적용 방법

본 절에서는 웹 동영상 페이지의 타입에 따라 키워드를 추출 방법을 제안한다. 그림 3은 웹 동영상 페이지로부터 키워드 추출 과정을 나타낸 그림이다.

첫번째 단계로 웹 페이지에 포함된 동영상의 수와 주변 텍스트 분포 상태를 바탕으로 하여 타입을 결정한 후, 다음 단계로 타입마다 다른 추출 알고리즘을 사용하여 동영상과 관련된 텍스트를 추출한다. 추출한 텍스트로부터 단어의 중요도에 따라 다른 가중치를 적용하는

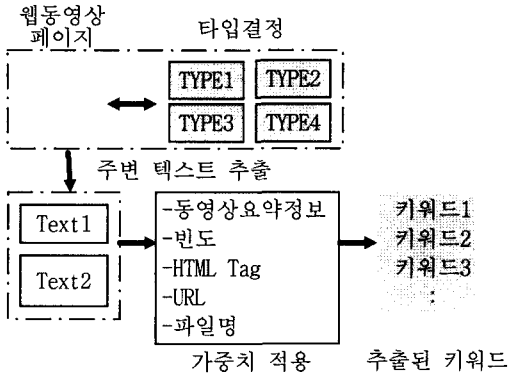


그림 3 키워드 추출 과정

데, 기존의 웹 이미지 페이지의 키워드 추출[1-3]에서 사용한 가중치 대상인 빈도수나 HTML 태그 정보 외에 추가적으로 웹 동영상 페이지의 의미적 특성인 동영상 상세요약정보의 포함여부를 적용한다. 텍스트 블록이 가지는 가중치( $W_{tb}$ )와 단어가 가지는 가중치( $W_{word}$ )로 웹 페이지내 포함되는 단어들의 전체 중요도( $W_{Total}$ )를 구하는 방법은 다음과 같다.

$$\begin{aligned}
 W_{Total} &= W_{tb} * W_{word} \\
 W_{tb} &= 1/d(N_v, N_t) \\
 W_{word} &= W_{Detail} * W_{TF} * W_{Tag}
 \end{aligned}
 \tag{3}$$

텍스트 블록의 가중치는 레이아웃 거리에 반비례하고,  $W_{Detail}$ 은 상세정보에 포함된 단어의 중요도이며,  $W_{TF}$ 는 단어의 빈도수,  $W_{Tag}$ 는 HTML 태그에 의한 중요도를 의미한다.

### 3.3.2 키워드 추출 알고리즘

웹 페이지로부터 키워드를 추출하는 단계는 앞 절에서 결정된 타입과 추출된 동영상 주변 텍스트정보로부터 결정된다. 그림 4에 웹 페이지에서 키워드를 추출하는 알고리즘을 나타내었다. ExtractKeyword()함수의 입력으로는 동영상을 가진 웹 페이지(WebVideoPage)와 추출할 키워드 개수( $n$ )가 주어진다. 먼저 웹 페이지에서 동영상과 주변 텍스트 블록을 추출한 후, 웹 페이지의 타입에 따라 가중치를 조정한다. TYPE1의 경우 추출된 텍스트에 모두 동일한 중요도를 적용하며, TYPE2는 레이아웃 거리에 따라 가중치를 조정한다. TYPE3의 경우 동영상에 여러 개 포함 되어있기 때문에 주변 텍스트가 어떤 동영상과 상관관계가 있는지를 결정해주어야 한다. TYPE4의 경우에는 동영상으로부터 떨어진 거리와 텍스트가 어떤 동영상과 관계가 있는지를 동시에 고려한다. 끝으로 단어들의 빈도수와 HTML 태그, 동영상 상세정보 등의 가중치를 반영하는 GetSpecialWeight()함수를 호출하여 최종적으로 가중치를 결정하고, 상위  $n$ 개의 키워드를 돌려준다.

```

//WebVideoPage : 웹 동영상 페이지
//n : 추출할 키워드 개수
//Type_w : 웹 페이지 타입
//V : 웹 페이지에서 추출한 동영상
//STB : 동영상 주변 텍스트 블록
//extractedkwd : 추출된 키워드 대상 단어 목록
Algorithm ExtractKeywords(WebVideoPage, n){
    Type_w = GetWebPageType(WebVideoPage);
    V = GetVideo(WebVideoPage);
    STB = GetSurroundTextBlocks(WebVideoPage);
    extractedkwd = {};
    for  $\forall v_i (v_i \in V)$  {
        for  $\forall tb_j (tb_j \in STB)$  {
            switch(Type_w) {
                case 1:
                     $W_j = 1;$ 
                    break;
                case 2:
                     $W_j = 1/d(v_i, tb_j);$ 
                    break;
                case 3:
                    check if  $tb_j$  is related to  $v_i$ ;
                    if (related)  $W_j = 1;$ 
                    else continue;
                    break;
                case 4:
                    check if  $tb_j$  is related to  $v_i$ ;
                    if (related)  $W_j = 1/d(v_i, tb_j);$ 
                    else continue;
                    break;
            }
            for  $\forall wd_k (wd_k \in tb_j)$  {
                 $W_k = W_j * GetSpecialWeight(wd_k);$ 
                extractedkwd = extractedkwd  $\cup$  { $wd_k$ };
            }
        }
    }
    return high position  $n$  keywords from extractedkwd;
}
    
```

그림 4 키워드 추출 알고리즘

## 4. 실험 및 결과 분석

본 장에서는 제안한 방법을 기반으로 하여 구현한 웹 동영상 키워드 추출 시스템에 대하여 설명하고, 실험을 통한 결과를 분석하여 제안한 방법의 성능을 평가한다.

### 4.1 성능평가 방법 및 데이터

본 논문에서 제안한 방법에 의한 키워드 추출 결과를 평가하고, 기존의 방법과의 비교를 위해서 사람에 의해 선택된 웹 동영상에 대한 키워드와 제안한 방법 및 기존의 방법에 의해 추출된 키워드를 비교한다. 기존의 방법으로는 대표적인 웹 이미지 검색 시스템에서 사용된 키워드 추출 방법 중 한가지인 ImageRover[2] 방식을 이용하였다. 정보검색시스템의 성능을 평가하기 위한 척도로는 재현율, 정확율, 만족율, 적용율 등이 있으나 본 논문에서는 재현율을 평가적으로 사용한다. 실험을 위해 사용한 데이터는 본 논문에서 구분한 웹 동영상 페이지의 4가지 타입 각각 30개로 이루어졌다. 실험에 사용한 식 (3)의 가중치는 표 2와 같다.

표 2 실험에 적용한 가중치

분류	항목	가중치
$W_{Tag}$	<b>, <i>	2.0
	<font size>	1) $Size_{Avg} = Size_{Font} : 1.0$ 2) $Size_{Avg} < Size_{Font} \leq (Size_{Avg} + 2) : 2.0^1$ 3) $(Size_{Avg} + 2) < Size_{Font} : 2.0^2$ 4) $(Size_{Avg} - 2) \leq Size_{Font} < Size_{Avg} : 2.0^{-1}$ 5) $Size_{Font} < (Size_{Avg} - 4) : 2.0^{-2}$
$W_{Detail}$	모든항목	4.0
$W_{TF}$	-	1.0 * 빈도수

사람에 의해 추출된 키워드와 본 논문에서 제안한 방법에 의한 키워드 추출 결과 사이의 재현율을 구하고, 마찬가지로 기존 웹 이미지 키워드 추출 방법에 의한 키워드 추출 결과에 대하여 재현율을 구한다. 두 결과를 비교함으로써 제안한 방법이 기존의 방법 보다 얼마나 성능향상을 보이는지를 설명한다. 타입 분류에 있어서는 T의 값을 몇으로 정하는지에 따라 성능이 달라 질 수 있다. 따라서 본 논문에서는 실험을 통해서 T의 값을 1~4로 바꾸어 가면서 재현율을 측정하였고, 실험 결과 T=1일 때, 가장 우수한 성능을 보였기 때문에, 본 실험에서는 T=1인 경우에 대하여, 웹 페이지의 타입을 분류하고, 타입 별 키워드 추출 성능을 측정한다.

4.2 결과분석

T의 값이 1일 때, 제안한 방법과 ImageRover[2]와 성능을 비교해 본 결과는 그림 5와 같다. 먼저 그림 5(a)는 TYPE1에 대한 재현율을 나타내는데 키워드 개수가 4~5개일 때 재현율이 ImageRover에 비해 약간 떨어지게 나타났다. 이는 ImageRover가 TYPE1과 같이 간단한 구조의 웹 페이지를 대상으로 하고 있기 때문에 제안한 방법에 대한 성능이 ImageRover와 거의 동일하거나 약간 낮게 나왔다고 볼 수 있다. 웹 페이지의 구조가 TYPE2에 속할 경우에 대한 성능은 그림 5(b)에 나타난다. 추출 키워드가 많아질수록 제안한 방법이 ImageRover에 비해 뛰어난 성능을 나타냄을 알 수 있다. 이는 ImageRover가 웹 페이지의 구조적인 특성은 고려하지 않고, 텍스트 상의 거리만을 평가하여 키워드를 추출했기 때문이다. 이에 비해서 본 논문에서 제안한 방법은 복잡한 구조의 페이지에서의 동영상과 텍스트의 관계를 평가하여 키워드를 추출하기 때문에 ImageRover에 비하여 높은 성능을 낸다고 할 수 있다.

다음으로 그림 5(c)는 TYPE3에 속하는 웹 동영상 페이지에 대한 성능을 나타낸다. TYPE3는 다수개의 동영상을 포함하는 간단한 구조의 웹 동영상 페이지로 구성되어 있어서 추출 과정에서 다른 동영상과 관련 있는 텍스트라고 판단되는 부분에서는 키워드를 추출 하지 않게 된다. 따라서 추출 되어야 하는 키워드가 의도하지

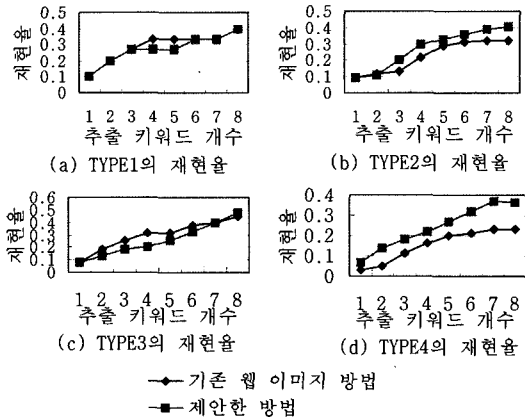


그림 5 타입별 실험결과(재현율)

않게 필터링 되는 경우가 생길 수 있다. 따라서 이러한 이유로 TYPE3에 대한 키워드 추출 재현율이 추출 개수 7이하에 대하여 ImageRover보다 조금 낮게 나타난다고 할 수 있다. 마지막으로 TYPE4에 대한 성능은 그림 5(d)와 같다. TYPE4는 페이지 내 포함된 동영상의 수도 많고, 구성방식도 복잡하다. 따라서 ImageRover는 이러한 특성들을 고려하지 않기 때문에 재현율이 제안한 방법에 비하여 전체적으로 떨어지는 것을 볼 수 있다.

제안한 방법이 특정 타입에 대한 적용이 아니라 일반적인 경우에도 적용하는 것이 가능한지를 알아보기 위해서 전체 타입에 종합적인 성능을 비교하였다. 그림 6은 전체 120개의 페이지에 대한 재현율을 나타내는데, 추출한 키워드의 개수가 5개일 때까지는 ImageRover와 유사한 재현율을 보이다가 추출 개수가 많아지면서 비교적 높은 재현율을 나타내고 있다. 따라서 본 논문에서 제안한 방법은 추출하는 키워드의 개수가 많은 경우에서 ImageRover에 비하여 올바른 키워드를 선택하는 능력이 뛰어남을 알 수 있다.

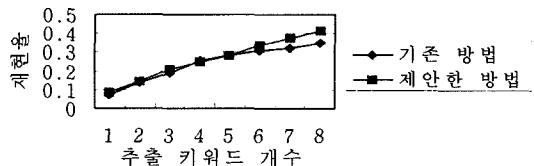


그림 6 전체에 대한 실험결과

실험결과 본 논문에서 제안한 방법이 ImageRover와 비교하여 전체적으로는 재현율에 대해 18%의 성능 향상(키워드 8개 기준)을 보였다. 그리고 타입별 실험을 통하여, 제안한 방법이 간단한 웹 페이지 구성 방법인 TYPE1의 경우를 제외하고는 모든 타입에서 ImageRover

와 비교하여 성능향상을 얻었으며, 특히 복잡한 형태의 웹 페이지 구조인 TYPE2, TYPE4에서는 ImageRover와 비교하였을 때 높은 성능평가를 보였다. 따라서 웹 동영상 페이지의 타입을 분류하고, 타입별로 키워드 추출방법을 달리하여 추출하는 것은 의미가 있다고 할 수 있다.

5. 결론 및 향후 연구 방향

웹 이미지와 동영상 페이지를 분석한 결과 웹 페이지에서 이미지와 동영상이 차지하는 의미가 다르고, 페이지 구성방식에도 차이를 보이기 때문에 웹 동영상 검색에서는 기존의 웹 이미지에서 사용한 방법과는 다른 방법이 요구된다. 따라서 본 논문에서는 웹 동영상 페이지의 동영상과 텍스트의 구성 방식을 분석하고, 페이지 구성에 영향을 주는 요소를 바탕으로 페이지 구조를 4가지 타입으로 분류하였다. 그리고 각 타입 별 키워드 추출 알고리즘을 제안하였다. 실험을 통하여 평가하여 본 결과, 본 논문에서 제안한 방법이 기존 웹 이미지 키워드 추출 방법에 비해서 실험대상(120개의 웹 동영상 페이지)에 대하여 18%의 (키워드 8개 기준) 성능 향상을 보였다. 또한 타입별 30개의 페이지에 대한 실험 결과, 제안한 방법이 TYPE1의 경우를 제외하고는 모든 타입에서 기존 방법과 비교하여 성능향상을 얻었으며, 특히 복잡한 형태의 웹 페이지 구조인 TYPE2, TYPE4에서는 기존 이미지 방법과 비교하였을 때 높은 성능 평가를 보였다. 따라서 웹 동영상 페이지의 타입을 분류하고, 타입별로 키워드 추출방법을 달리하여 추출하는 본 논문의 제안 방법은 웹에 분포한 일반적인 웹 동영상 페이지로부터 키워드를 추출하는 방법으로 사용할 경우 기존의 방법에 비해 우수한 성능을 보일 것으로 예상된다.

본 논문에서 제안한 방법은 웹 동영상 관련 키워드 추출 방법에 대한 초기 연구에 해당하며, 따라서 개선해야 부분들을 가지고 있다. 먼저 본 논문은 제한적인 웹 동영상 페이지에 대한 분석을 기반으로 출발하였다. 모든 웹 동영상 페이지의 분석을 바탕으로 한 것이 아니기 때문에 실제 임의의 웹 동영상 페이지로부터 키워드를 추출할 경우 본 논문에서 실험한 결과와는 어느 정도 차이를 보일 수 있다. 또한 본 논문은 웹 동영상 페이지의 텍스트 정보만을 바탕으로 키워드를 추출하고 있으나 동영상의 내용정보를 함께 이용하여 키워드 추출의 정확도를 높이기 위한 연구의 진행이 요구된다.

참 고 문 헌

[1] C. Frankel, M. J. Swain, and V. Athitsos, *WebSeer: An Image Search Engine for the World Wide Web*, Technical Report 96-14, University of Chi-

cago Computer Science Department, 1996.  
 [2] S. Sclaroff, L. Taycher and M. L. Cascia, "Image-Rover: A Content-Based Image Browser for the World Wide Web," in *Proceedings of IEEE Workshop on Content-based Access of Image and Video Libraries*, pp. 2-9, 1997.  
 [3] S. Mukerjea and J. Cho, "Automatically Determining Semantics for World Wide Web Multimedia Information Retrieval," in *Journal of Visual Languages and Computing*, Vol.10, pp. 585-606, 1999.  
 [4] M. Costa, N. Correia and N. Guimaraes, "Annotations as Multiple Perspectives of Video Content," in *Proceedings of the 10th ACM International conference on Multimedia*, pp. 283-286, 2002.  
 [5] T. Volkmer, J. R. Smith, and A. Natsev, "A Web-based System for Collaborative Annotation of Large Image and Video Collections," in *Proceedings of the 13th annual ACM International Conference on Multimedia*, pp. 892-901, 2005.  
 [6] M. Bertini, A. Bimbo and C. Torniai, "Automatic Video Annotation using Ontologies Extended with Visual Information," in *Proceedings of the 13th annual ACM International Conference on Multimedia*, pp. 395-398, 2005.  
 [7] S. Rui, W. Jin, and T. Shua, "A Novel Approach to Auto Image Annotation Based on Pair-wise Constrained Clustering and Semi-naive Bayesian Model," in *Proceedings of IEEE 11th International Conference on Multimedia Modeling*, pp. 322-327, 2005.  
 [8] J. Z. Wang and J. Li, "Learning-based Linguistic Indexing of Pictures with 2-D MHMMs," in *Proceedings of the 10th ACM International Conference on Multimedia*, pp. 436-445, 2002.  
 [9] D. Cai, S. Yu, J. R. Wen and W. Y. Ma, *VIPS:a Vision-based Page Segmentation Algorithm*, MSR-TR-2003-79/Microsoft Research, 2003.



이 중 원  
 1987년 서강대학교 전자계산학과(학사)  
 1996년 서강대학교 공공정책대학원 정보처리학과(석사). 2003년 서강대학교 대학원 컴퓨터학과(박사수료). 현재 청강문화산업대학 e스포츠게임과 부교수. 관심분야는 멀티미디어검색, 웹 마이닝, 서버

및 보안



최 기 석  
 2006년 서강대학교 컴퓨터학과(학사). 2008년 서강대학교 대학원 컴퓨터학과(석사) 현재 서강대학교 대학원 박사과정 입학 관심분야는 멀티미디어시스템, 멀티미디어 검색, 시맨틱 웹



장 주 연

2005년 서강대학교 컴퓨터학과(학사). 2007년 서강대학교 대학원 컴퓨터학과(석사)  
현재 삼성전자 디지털미디어 사업부. 관심분야는 멀티미디어시스템, 웹 동영상 검색



남 종 호

1986년 2월 서강대학교 전자계산학과 졸업(학사). 1988년 2월 한국과학기술원 전산과 졸업(석사). 1992년 2월 한국과학기술원 전산과 졸업(박사). 1992년 3월~1992년 8월 한국과학기술원 정보전자 연구소(연구원). 1992년 9월~1993년 8월 일본 Fujitsu 연구소(방문연구원). 1993년 9월~현재 서강대학교 컴퓨터학과 교수. 관심분야는 멀티미디어 시스템, 동영상 검색, 동영상 분석, 병렬/분산 처리, 인터넷 컴퓨팅